S-AMP for Non-linear Observation Models

Cakmak, Burak; Winther, Ole; Fleury, Bernard Henri

Published in:
IEEE International Symposium on Information Theory (ISIT), 2015

DOI (link to publication from Publisher):
10.1109/ISIT.2015.7282968

Publication date:
2015

Document Version
Early version, also known as pre-print

S-AMP for Non-linear Observation Models

Burak Çakmak
Department of Electronic Systems
Aalborg University
9220 Aalborg, Denmark
Email: buc@es.aau.dk

Ole Winther
DTU Compute
Technical University of Denmark
2800 Lyngby, Denmark
Email: olwi@dtu.dk

Bernard H. Fleury
Department of Electronic Systems
Aalborg University
9220 Aalborg, Denmark
Email: fleury@es.aau.dk

***Abstract*—Recently we presented the S-AMP approach, an extension of approximate message passing (AMP), to be able to handle general invariant matrix ensembles. In this contribution we extend S-AMP to non-linear observation models. We obtain generalized AMP (GAMP) as the special case when the measurement matrix has zero-mean iid Gaussian entries. Our derivation is based upon 1) deriving expectation-propagation-(EP)-like equations from the stationary-points equations of the Gibbs free energy under first- and second-moment constraints and 2) applying additive free convolution in free probability theory to get low-complexity updates for the second moment quantities.**

***Index Terms*—Approximate Message Passing, Variational Inference, Expectation Propagation, Free Probability**

I. Introduction

Approximate message passing techniques, e.g. [1]–[3], have recently received significant attention by the signal processing community. Essentially, these methods are based on taking the large system limit of loopy belief propagation where the central limit theorem can be applied when the underlying measurement matrix has independent and zero-mean entries.

Variational inference techniques are well-established in the field of information theory e.g. [4], [5] and machine learning e.g. [6], [7]. For example, it is well-known that exact inference can be formulated as the solution to a minimization problem of the Gibbs free energy of the underlying probabilistic model under certain marginalization consistency constraints [4]. We have recently shown in [8] that for the zero-mean independent identically distributed (iid) measurement matrix, approximate message passing (AMP) algorithm [1] can also be obtained from the stationary-points equations of the Gibbs energy under first- and second-moment consistency constraints Furthermore, AMP can be extended to general *invariant*[1] matrix ensembles by means of the asymptotic spectrum of the measurement matrix. We call this approach S-AMP (where S comes from the fact that the derivation uses the S-transform).

In practice there are many interesting cases where the observation model is non-linear, e.g non-linear form of compressed sensing, Gaussian processes for classification. In this article we extend the S-AMP approach [8] to general observation models. Specifically we address the sum-product generalized AMP (GAMP for short) algorithm [3].

[1]Note that we omit to mention the invariance property in [8]. It is however crucial for the derivation.

The derivation of GAMP is based on certain approximations (mainly Gaussian and quadratic approximations) of loopy belief propagation. If the measurement matrix is large and has zero mean and iid entries, GAMP provides excellent performance, e.g. [3], [9]. Furthermore, for general matrix ensembles it can show quite reasonable accuracy [10]. However the algorithm itself and its derivation are not well-understood.

To better understand GAMP, in [11] the authors characterize its fixed points. Specifically, they show that GAMP can be obtained from the stationary-point equations of some implicit approximations of naive mean-field approximation [11]. These implicit approximations only provide limited insight. Furthermore, the naive mean-field interpretation is misleading, because the fixed points of AMP-type algorithms are typically known as the TAP-like equations, i.e. they include a correction term to naive mean-field solution. In fact GAMP can also be obtained from the stationary-points equations of the Bethe free energy (BFE) of the underlying loopy graph under first- and second-moment constraints. However, this approach also limits our understanding, because the BFE formulation of a loopy graph is suitable for sparsely connected systems.

In this work we focus on the BFE formulation of a tree graph, i.e. an exact Gibbs free energy formulation. We note that our approach coincides with expectation propagation (EP) [12]–[14] since the fixed points of EP are the stationary points of BFE of the underlying probabilistic graph under a set of moment consistency constraints [6].

Notations: The entries of the $N \times K$ matrix $\boldsymbol{X}$ are denoted by either X_{nk} or $[X]_{nk}$, $n \in \mathcal{N} \triangleq \{n : 1 \leq n \leq N\}$ and $k \in \mathcal{K} \triangleq \{k : 1 \leq k \leq K\}$. The transposition is denoted by $(\cdot)^{\dagger}$. The entries of a vector $\boldsymbol{u} \in \mathbb{R}^{T\times 1}$ are indicated by either u_t or $[u]_t$, $1 \leq t \leq T$. Furthermore $\langle \boldsymbol{u} \rangle \triangleq \sum_{t=1}^{T} u_t/T$. Moreover, $\mathrm{diag}(\boldsymbol{u})$ is a diagonal matrix with the elements of vector $\boldsymbol{u}$ on the main diagonal. For a square matrix $\boldsymbol{X}$, $\mathrm{diag}(\boldsymbol{X})$ is a column vector containing the diagonal elements of $\boldsymbol{X}$. Furthermore $\mathrm{Diag}(\boldsymbol{X}) \triangleq \mathrm{diag}(\mathrm{diag}(\boldsymbol{X}))$. We denote by $\Re z$ and $\Im z$ the real and imaginary parts of $z \in \mathbb{C}$, respectively. The Gaussian probability density function (pdf) with mean $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ is denoted by $N(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$. Throughout the paper when referring to "in the large system limit" we imply that N, K tends to infinity with the ratio $\alpha \triangleq N/K$ fixed. All large system limits are assumed to hold in the almost sure sense, unless explicitly stated.

II. System Model and Review of GAMP

Consider the estimation of a random vector $\boldsymbol{x} \in \mathbb{R}^{K\times 1}$ which is linearly transformed by $\boldsymbol{A} \in \mathbb{R}^{N\times K}$ as $\boldsymbol{z} \triangleq \boldsymbol{A}\boldsymbol{x}$, then passed through a noisy channel whose output is given by $\boldsymbol{y} \in \mathbb{R}^{N\times 1}$. We assume that the conditional pdf of the channel factorizes according to

$$p(\boldsymbol{y}|\boldsymbol{z}) = \prod_{n\in\mathcal{N}} p(y_n|z_n). \tag{1}$$

Furthermore for the Bayesian setting we assign a prior pdf for $\boldsymbol{x}$ that is assumed to factorizes as

$$p(\boldsymbol{x}) = \prod_{k\in\mathcal{K}} p_k(x_k). \tag{2}$$

A. GAMP summarized

We summarize GAMP here for the sake of streamlining and making the connection to the extension of S-AMP. We separate the GAMP iteration rules [3] into two parts: (i) GAMP-1st order that initializes $\hat{\boldsymbol{x}}^t$, $\boldsymbol{\tau}_{\rm x}^t$ and $\boldsymbol{m}^t$ from tabula rasa at $t \leq 0$ and proceeds iteratively as

$$\boldsymbol{\kappa}_{\rm z}^t = \boldsymbol{A}\hat{\boldsymbol{x}}^t - (\mathbf{V}_{\rm z}^t)^{-1}\boldsymbol{m}^{t-1} \tag{3}$$
$$\hat{\boldsymbol{z}}^t = \mu_{\rm z}(\boldsymbol{\kappa}_{\rm z}^t; \mathbf{V}_{\rm z}^t) \tag{4}$$
$$\boldsymbol{\tau}_{\rm z}^t = \sigma_{\rm z}(\boldsymbol{\kappa}_{\rm z}^t; \mathbf{V}_{\rm z}^t) \tag{5}$$
$$\boldsymbol{m}^t = \mathbf{V}_{\rm z}^t(\hat{\boldsymbol{z}}^t - \boldsymbol{\kappa}_{\rm z}^t) \tag{6}$$
$$\boldsymbol{\kappa}_{\rm x}^t = (\mathbf{V}_{\rm x}^t)^{-1}\boldsymbol{A}^\dagger\boldsymbol{m}^t + \hat{\boldsymbol{x}}^t \tag{7}$$
$$\hat{\boldsymbol{x}}^{t+1} = \mu_{\rm x}(\boldsymbol{\kappa}_{\rm x}^t; \mathbf{V}_{\rm x}^t) \tag{8}$$
$$\boldsymbol{\tau}_{\rm x}^{t+1} = \sigma_{\rm x}(\boldsymbol{\kappa}_{\rm x}^t; \mathbf{V}_{\rm x}^t). \tag{9}$$

(ii) GAMP-2nd order are the update rules for $\mathbf{V}_{\rm z}^t$ and $\mathbf{V}_{\rm x}^t$:

$$\mathbf{V}_{\rm z}^t = (\mathrm{diag}((\boldsymbol{A}\circ\boldsymbol{A})\boldsymbol{\tau}_{\rm x}^t))^{-1} \tag{10}$$
$$\boldsymbol{\tau}_{\rm m}^t = \mathbf{V}_{\rm z}^t(\mathbf{1} - \mathbf{V}_{\rm z}^t\boldsymbol{\tau}_{\rm z}^t) \tag{11}$$
$$\mathbf{V}_{\rm x}^t = \mathrm{diag}((\boldsymbol{A}\circ\boldsymbol{A})^\dagger\boldsymbol{\tau}_{\rm m}^t). \tag{12}$$

In these expressions $\mathbf{1}$ is the all-ones vector of appropriate dimension and $\mu_{\rm x}$ and $\sigma_{\rm x}$ are scalar functions. Specifically, if $\mathbf{V}$ is a $K\times K$ diagonal matrix and $\boldsymbol{\kappa}$ is a $K\times 1$ vector, then for $k\in\mathcal{K}$ $[\mu_{\rm x}(\boldsymbol{\kappa};\mathbf{V})]_k$ and $[\sigma_{\rm x}(\boldsymbol{\kappa};\mathbf{V})]_k$ are respectively the mean and the variance of the pdf

$$q_k(x_k) \propto p_k(x_k)\exp\left(-\frac{\mathrm{V}_{kk}}{2}(x_k-\kappa_k)^2\right). \tag{13}$$

Similarly, $\mu_{\rm z}$ and $\sigma_{\rm z}$ are scalar functions such that if $\mathbf{V}$ is a $N\times N$ diagonal matrix and $\boldsymbol{\kappa}$ is a $N\times 1$ vector, for $n\in\mathcal{N}$ $[\mu_{\rm z}(\boldsymbol{\kappa};\mathbf{V})]_n$ and $[\sigma_{\rm z}(\boldsymbol{\kappa};\mathbf{V})]_n$ are respectively the mean and the variance of the pdf

$$q_n(z_n) \propto p(y_n|z_n)\exp\left(-\frac{\mathrm{V}_{nn}}{2}(z_n-\kappa_n)^2\right). \tag{14}$$

If the entries of $\boldsymbol{A}$ are iid with zero mean and variance $1/N$, the iteration steps for the GAMP-2nd order simplify as

$$\mathbf{V}_{\rm z}^t = \frac{\alpha}{\langle\boldsymbol{\tau}_{\rm x}^t\rangle}\mathbf{I}, \quad \mathbf{V}_{\rm x}^t = \langle\boldsymbol{\tau}_{\rm m}^t\rangle\mathbf{I}, \tag{15}$$

where $\mathbf{I}$ is the identity matrix of appropriate dimension. We note that if in addition $p(\boldsymbol{y}|\boldsymbol{z}) = N(\boldsymbol{y};\boldsymbol{z},\sigma^2\mathbf{I})$, GAMP yields AMP, see e.g. [2, Appendix C].

III. Gibbs Free Energy with Moment Constraints

For the sake of notational compactness, consider $\boldsymbol{s} = (\boldsymbol{x},\boldsymbol{z})$. Furthermore we introduce the set $\mathcal{V} \triangleq \mathcal{K}\cup\mathcal{N}$ and assume that $\mathcal{K}$ and $\mathcal{N}$ are disjoint. Moreover we define

$$f_A(\boldsymbol{s}) \triangleq \delta(\boldsymbol{z} - \boldsymbol{A}\boldsymbol{x}) \tag{16}$$
$$f_v(s_v) \triangleq \begin{cases} p_v(x_v) & v\in\mathcal{K} \\ p(y_v|z_v) & v\in\mathcal{N}. \end{cases} \tag{17}$$

With these definitions, the posterior pdf of $\boldsymbol{s}$ reads

$$p(\boldsymbol{s}|\boldsymbol{y},\boldsymbol{A}) = \frac{1}{Z}f_A(\boldsymbol{s})\prod_{v\in\mathcal{V}} f_v(s_v). \tag{18}$$

with Z denoting a normalization constant. The factor graph representing (18) is a tree. Thus the BFE of (18) is equal to its Gibbs free energy [4]:

$$\begin{aligned}\mathrm{G}(\{\tilde{b}_v, b_A, b_v\}) \triangleq &-\sum_{v\in\mathcal{V}}\int \tilde{b}_v(s_v)\log\tilde{b}_v(s_v)\mathrm{d}s_v \\ &-\int b_A(\boldsymbol{s})\log\frac{f_A(\boldsymbol{s})}{b_A(\boldsymbol{s})}\mathrm{d}\boldsymbol{s} - \sum_{v\in\mathcal{V}}\int b_v(s_v)\log\frac{f_v(s_v)}{b_v(s_v)}\mathrm{d}s_v.\end{aligned} \tag{19}$$

Here b_A and b_v, $v\in\mathcal{V}$, denote the beliefs of the factors in (18), while $\tilde{b}_v$, $v\in\mathcal{V}$, denote the beliefs of the unknown variables in (18). Without loss of generality we assume that the expressions $f_A(\boldsymbol{s})/b_A(\boldsymbol{s})$ and $f_v(s_v)/b_v(s_v)$ in (19) are strictly continuous; so that the Gibbs free energy is well-defined. Indeed this is what we will end up with in the analysis.

If we define a Lagrangian for (19) that accounts for certain marginalization consistency constrains, then at its stationary point, the belief $\tilde{b}_v(s_v)$ is equal to $p(s_v|\boldsymbol{y},\boldsymbol{A})$ for all $v\in\mathcal{V}$ [4]. Instead, following the arguments of [6], we define the Lagrangian on the basis of a set of moment consistency constraints as

$$\begin{aligned}\mathcal{L}(\{\tilde{b}_v, b_A, b_v\}) \triangleq &\ \mathrm{G}(\{b_v, b_A, \tilde{b}_v\}) + \mathcal{Z} \\ &-\sum_{v\in\mathcal{V}}\boldsymbol{\nu}_v^\dagger\int\boldsymbol{\phi}(s_v)\left\{b_A(\boldsymbol{s}) - \tilde{b}_v(s_v)\right\}\mathrm{d}\boldsymbol{s} \\ &-\sum_{v\in\mathcal{V}}\bar{\boldsymbol{\nu}}_v^\dagger\int\boldsymbol{\phi}(s_v)\left\{b_v(s_v) - \tilde{b}_v(s_v)\right\}\mathrm{d}s_v.\end{aligned} \tag{20}$$

Here we consider constraints on the mean and variance, i.e. $\boldsymbol{\phi}(s_v) = (s_v, s_v^2)$, $v\in\mathcal{V}$. For convenience we write the Lagrangian multipliers as

$$\boldsymbol{\nu}_v \triangleq \left(\gamma_v, -\frac{\Lambda_{vv}}{2}\right), \quad \bar{\boldsymbol{\nu}}_v \triangleq \left(\rho_v, -\frac{\mathrm{V}_{vv}}{2}\right), \quad v\in\mathcal{V}.$$

The term $\mathcal{Z}$ in (20) accounts for the normalization constraints:

$$\begin{aligned}\mathcal{Z} \triangleq &-\beta_A\left(1 - \int b_A(\boldsymbol{s})\mathrm{d}\boldsymbol{s}\right) \\ &-\sum_{v\in\mathcal{V}}\tilde{\beta}_v\left(1 - \int\tilde{b}_v(s_v)\mathrm{d}s_v\right) - \beta_v\left(1 - \int b_v(s_v)\mathrm{d}s_v\right)\end{aligned}$$

where β_A, β_v, $\tilde{\beta}_v$ are the associated Lagrange multipliers.

A. The Stationary Points of the Lagrangian

We formulate the estimate of $s_v, v \in \mathcal{V}$, as

$$\hat{s}_v \triangleq \int s_v \tilde{b}_v^\star(s_v) \mathrm{d}s_v, \tag{21}$$

where $\tilde{b}_v^\star(s_v)$ represents $\tilde{b}_v(s_v)$ at a stationary point of the Lagrangian (20).

For notational convenience we introduce first the $(K+N)\times(K+N)$ diagonal matrices $\mathbf{\Lambda}$ and $\mathbf{V}$ as well as the $(K+N)\times 1$ vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\rho}$ whose entries are respectively Λ_{vv}, V_{vv}, γ_v and ρ_v, $v \in \mathcal{V}$. In connection with variables $\boldsymbol{x}$ and $\boldsymbol{z}$ we write

$$\mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}_{\mathrm{x}} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_{\mathrm{z}} \end{pmatrix}, \quad \boldsymbol{\gamma} = (\boldsymbol{\gamma}_{\mathrm{x}}, \boldsymbol{\gamma}_{\mathrm{z}}) \tag{22}$$

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{\mathrm{x}} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{\mathrm{z}} \end{pmatrix}, \quad \boldsymbol{\rho} = (\boldsymbol{\rho}_{\mathrm{x}}, \boldsymbol{\rho}_{\mathrm{z}}). \tag{23}$$

The dimensions of $\mathbf{\Lambda}_{\mathrm{x}}$ and $\mathbf{V}_{\mathrm{x}}$ are $K \times K$; vectors $\boldsymbol{\gamma}_{\mathrm{x}}$ and $\boldsymbol{\rho}_{\mathrm{x}}$ have dimension $K \times 1$.

Following the arguments of [6], the stationary points of the Lagrangian (20) are given by

$$\tilde{b}_v^\star(s_v) = \frac{1}{\tilde{Z}_v} \exp((\boldsymbol{\nu}_v + \bar{\boldsymbol{\nu}}_v)^\dagger \boldsymbol{\phi}(s_v)), \quad v \in \mathcal{V} \tag{24}$$

$$b_v^\star(s_v) = \frac{1}{Z_v} f_v(s_v) \exp(\bar{\boldsymbol{\nu}}_v^\dagger \boldsymbol{\phi}(s_v)), \quad v \in \mathcal{V} \tag{25}$$

$$b_A^\star(\boldsymbol{s}) = \frac{1}{Z_A} f_A(\boldsymbol{s}) \exp\left(-\frac{1}{2}\boldsymbol{s}^\dagger \mathbf{\Lambda} \boldsymbol{s} + \boldsymbol{s}^\dagger \boldsymbol{\gamma}\right) \tag{26}$$

where Z_A, Z_v, $\tilde{Z}_v$ are the associated normalization constants.

Let us first consider the marginalization of the belief $b_A^\star(\boldsymbol{s})$ with respect to $\boldsymbol{z}$:

$$b_A^\star(\boldsymbol{x}) = \int b_A^\star(\boldsymbol{x}, \boldsymbol{z}) \mathrm{d}\boldsymbol{z} = N(\boldsymbol{x}; \hat{\boldsymbol{x}}, \mathbf{\Sigma}_{\mathrm{x}}) \tag{27}$$

where

$$\mathbf{\Sigma}_{\mathrm{x}} \triangleq (\mathbf{\Lambda}_{\mathrm{x}} + \boldsymbol{A}^\dagger \mathbf{\Lambda}_{\mathrm{z}} \boldsymbol{A})^{-1}, \quad \hat{\boldsymbol{x}} \triangleq \mathbf{\Sigma}_{\mathrm{x}}(\boldsymbol{\gamma}_{\mathrm{x}} + \boldsymbol{A}^\dagger \boldsymbol{\gamma}_{\mathrm{z}}). \tag{28}$$

Here we note that $\mathbf{\Sigma}_{\mathrm{x}}$ is positive definite since $b_A^\star(\boldsymbol{x})$ is a well-defined pdf. Let us then consider the marginalization over $\boldsymbol{x}$, which basically follows from the linear transformation property of a Gaussian random vector:

$$\begin{aligned} b_A^\star(\boldsymbol{z}) &= \frac{e^{-\frac{1}{2}\boldsymbol{z}^\dagger \mathbf{\Lambda}_{\mathrm{z}} \boldsymbol{z} + \boldsymbol{z}^\dagger \boldsymbol{\gamma}_{\mathrm{z}}}}{Z_A} \int \delta(\boldsymbol{z} - \boldsymbol{A}\boldsymbol{x}) e^{-\frac{1}{2}\boldsymbol{x}^\dagger \mathbf{\Lambda}_{\mathrm{x}} \boldsymbol{x} + \boldsymbol{x}^\dagger \boldsymbol{\gamma}_{\mathrm{x}}} \mathrm{d}\boldsymbol{x} \\ &= \int \delta(\boldsymbol{z} - \boldsymbol{A}\boldsymbol{x}) N(\boldsymbol{x}; \hat{\boldsymbol{x}}, \mathbf{\Sigma}) \mathrm{d}\boldsymbol{x} = N(\boldsymbol{z}; \hat{\boldsymbol{z}}, \mathbf{\Sigma}_{\mathrm{z}}) \end{aligned} \tag{29}$$

where $\hat{\boldsymbol{z}} \triangleq \boldsymbol{A}\hat{\boldsymbol{x}}$ and $\mathbf{\Sigma}_{\mathrm{z}} \triangleq \boldsymbol{A}\mathbf{\Sigma}_{\mathrm{x}}\boldsymbol{A}^\dagger$.

At this stage it is convenient to define

$$\boldsymbol{\kappa} \triangleq (\boldsymbol{\kappa}_{\mathrm{x}}, \boldsymbol{\kappa}_{\mathrm{z}}) = (\mathbf{V}_{\mathrm{x}}^{-1}\boldsymbol{\rho}_{\mathrm{x}}, \mathbf{V}_{\mathrm{z}}^{-1}\boldsymbol{\rho}_{\mathrm{z}}) \tag{30}$$

with $\boldsymbol{\kappa}_{\mathrm{x}} \in \mathbb{R}^K$. In this way we can write the belief in (25) as

$$b_v^\star(s_v) \propto f_v(s_v) \exp\left(-\frac{\mathrm{V}_{vv}}{2}(s_v - \kappa_v)^2\right), \quad v \in \mathcal{V}. \tag{31}$$

Thereby (31) has a form identical to (13) and (14) for $v \in \mathcal{N}$ and for $v \in \mathcal{K}$, respectively. Then let us define

$$\mu(\boldsymbol{\kappa}; \mathbf{V}) \triangleq (\mu_{\mathrm{x}}(\boldsymbol{\kappa}_{\mathrm{x}}; \mathbf{V}_{\mathrm{x}}), \mu_{\mathrm{z}}(\boldsymbol{\kappa}_{\mathrm{z}}; \mathbf{V}_{\mathrm{z}})) \tag{32}$$

$$\sigma(\boldsymbol{\kappa}; \mathbf{V}) \triangleq (\sigma_{\mathrm{x}}(\boldsymbol{\kappa}_{\mathrm{x}}; \mathbf{V}_{\mathrm{x}}), \sigma_{\mathrm{z}}(\boldsymbol{\kappa}_{\mathrm{z}}; \mathbf{V}_{\mathrm{z}})). \tag{33}$$

The entries $[\mu(\boldsymbol{\kappa}; \mathbf{V})]_v$ and $[\sigma(\boldsymbol{\kappa}; \mathbf{V})]_v$ are respectively the mean and variance of the belief (25). Moreover we introduce

$$\mathbf{\Sigma} \triangleq \begin{pmatrix} \mathbf{\Sigma}_{\mathrm{x}} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{\mathrm{z}} \end{pmatrix}, \quad \hat{\boldsymbol{s}} = (\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}}). \tag{34}$$

With these definitions, the identities resulting from the moment consistency constraints are given by

$$\hat{\boldsymbol{s}} = \mathrm{Diag}(\mathbf{\Sigma})(\boldsymbol{\gamma} + \boldsymbol{\rho}), \quad \hat{\boldsymbol{s}} = \mu(\boldsymbol{\kappa}; \mathbf{V}) \tag{35}$$

$$\mathrm{Diag}(\mathbf{\Sigma}) = (\mathbf{\Lambda} + \mathbf{V})^{-1}, \quad \mathrm{diag}(\mathbf{\Sigma}) = \sigma(\boldsymbol{\kappa}; \mathbf{V}). \tag{36}$$

B. The TAP-like Equations and GAMP-1st Order

By using the fixed-point identities presented in Section III-A, one can introduce numerous fixed-point algorithms. In this work we restrict our attention to TAP-like algorithms, e.g. [12], [14]. To that end we start with the definitions in (28) and write

$$\boldsymbol{\gamma}_{\mathrm{x}} = -\boldsymbol{A}^\dagger \boldsymbol{\gamma}_{\mathrm{z}} + (\mathbf{\Lambda}_{\mathrm{x}} + \boldsymbol{A}^\dagger \mathbf{\Lambda}_{\mathrm{z}} \boldsymbol{A})\hat{\boldsymbol{x}}. \tag{37}$$

Then, by making use of the identities in (35) and (36) we have

$$\begin{aligned} \boldsymbol{\rho}_{\mathrm{x}} &= \boldsymbol{A}^\dagger \boldsymbol{\gamma}_{\mathrm{z}} - (\mathbf{\Lambda}_{\mathrm{x}} + \boldsymbol{A}^\dagger \mathbf{\Lambda}_{\mathrm{z}} \boldsymbol{A})\hat{\boldsymbol{x}} + (\mathbf{\Lambda}_{\mathrm{x}} + \mathbf{V}_{\mathrm{x}})\hat{\boldsymbol{x}} && (38) \\ &= \boldsymbol{A}^\dagger(\boldsymbol{\gamma}_{\mathrm{z}} - \mathbf{\Lambda}_{\mathrm{z}} \boldsymbol{A}\hat{\boldsymbol{x}}) + \mathbf{V}_{\mathrm{x}}\hat{\boldsymbol{x}} && (39) \\ &= \boldsymbol{A}^\dagger \boldsymbol{m} + \mathbf{V}_{\mathrm{x}}\hat{\boldsymbol{x}} \quad \text{with} \quad \boldsymbol{m} \triangleq (\boldsymbol{\gamma}_{\mathrm{z}} - \mathbf{\Lambda}_{\mathrm{z}} \boldsymbol{A}\hat{\boldsymbol{x}}). && (40) \end{aligned}$$

Moreover, by the definition of $\boldsymbol{m}$ we also point out that

$$\begin{aligned} \boldsymbol{m} &= (\mathbf{\Lambda}_{\mathrm{z}} + \mathbf{V}_{\mathrm{z}})\hat{\boldsymbol{z}} - \boldsymbol{\rho}_{\mathrm{z}} - \mathbf{\Lambda}_{\mathrm{z}} \boldsymbol{A}\hat{\boldsymbol{x}} && (41) \\ &= \mathbf{V}_{\mathrm{z}}\hat{\boldsymbol{z}} - \boldsymbol{\rho}_{\mathrm{z}} = \mathbf{V}_{\mathrm{z}}(\hat{\boldsymbol{z}} - \boldsymbol{\kappa}_{\mathrm{z}}). && (42) \end{aligned}$$

Thereby we exactly obtain the fixed-point equations of GAMP-1st order, i.e. (3)-(9).

Now let us keep the iterations step of GAMP-1st order but define the update rule for $\mathbf{V}_{\mathrm{x}}^t$ and $\mathbf{V}_{\mathrm{z}}^t$ on the basis of the fixed-point identities in (36). For example:

$$\mathbf{\Lambda}_{\mathrm{z}}^t = (\mathrm{diag}(\boldsymbol{\tau}_{\mathrm{z}}^{t-1}))^{-1} - \mathbf{V}_{\mathrm{z}}^{t-1} \tag{43}$$

$$\mathbf{\Sigma}_{\mathrm{x}}^t = (\mathbf{\Lambda}_{\mathrm{x}}^{t-1} + \boldsymbol{A}^\dagger \mathbf{\Lambda}_{\mathrm{z}}^t \boldsymbol{A})^{-1} \tag{44}$$

$$\mathbf{V}_{\mathrm{z}}^t = \left(\mathrm{Diag}(\boldsymbol{A}\mathbf{\Sigma}_{\mathrm{x}}^t \boldsymbol{A}^\dagger)\right)^{-1} - \mathbf{\Lambda}_{\mathrm{z}}^t \tag{45}$$

$$\mathbf{\Lambda}_{\mathrm{x}}^t = (\mathrm{diag}(\boldsymbol{\tau}_{\mathrm{x}}^t))^{-1} - \mathbf{V}_{\mathrm{x}}^{t-1} \tag{46}$$

$$\mathbf{V}_{\mathrm{x}}^t = \left(\mathrm{Diag}(\mathbf{\Sigma}_{\mathrm{x}}^t)\right)^{-1} - \mathbf{\Lambda}_{\mathrm{x}}^t. \tag{47}$$

In this way we obtain a new fixed-point algorithm whose fixed points are the stationary point of Lagrangian (20). However from the complexity point of view these updates are problematic due to the matrix inversion in (44). In the sequel we will address how to circumvent this complexity problem as K, N are large.

C. The Large-System Simplifications

To bypass the need for matrix inversion such as (44) we utilize the so-called additive free convolution in free probability theory [15]. The reduction that we obtain in this way can be also obtained by means of the self-averaging ansatz in [14, Section 3.1].

In order to make use of additive free convolution we need to restrict our consideration to invariant matrix ensembles:

ASSUMPTION 1 *Consider the singular value decomposition $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}$ where $\boldsymbol{U}^{N\times N}$ and $\boldsymbol{V}^{K\times K}$ are orthogonal matrices and $\boldsymbol{D}$ is a $N \times K$ non-negative diagonal matrix. We distinguish between the invariance assumption on $\boldsymbol{A}$ from right and from left:* a) *$\boldsymbol{A}$ is invariant from right, i.e. $\boldsymbol{V}$ is Haar distributed;* b) *$\boldsymbol{A}$ is invariant from left, i.e. $\boldsymbol{U}$ is Haar distributed.*

It indeed makes sense to distinguish between the invariance from right and the invariance from left. For example, once we consider the classical linear observation model such as $p(\boldsymbol{y}|\boldsymbol{z}) = N(\boldsymbol{y}; \boldsymbol{z}, \sigma^2\mathbf{I})$, then $\boldsymbol{\Lambda}_{\mathrm{z}} = \mathbf{I}/\sigma^2$. In this case we do not need to consider Assumption 1-b).

We also make the following technical assumption on the limiting spectrum of the respective matrices:

ASSUMPTION 2 *As $N, K \to \infty$ with the ratio $\alpha = N/K$ fixed the spectra of $\boldsymbol{\Lambda}_{\mathrm{x}}$, $\boldsymbol{\Lambda}_{\mathrm{z}}$ and $\boldsymbol{A}^\dagger\boldsymbol{A}$ converge almost surely to some limiting spectra whose supports are compact.*

Due to lack of an explicit definition of the "Lagrangian" matrix $\boldsymbol{\Lambda}$, Assumption 2 is rather implicit. Nevertheless it can be considered in the same vein as the so-called thermodynamic limit in statistical physics: all microscopic variables converge to deterministic values in the thermodynamic limit [16].

Under Assumption 1-a) and Assumption 2, it turns out that $\boldsymbol{\Lambda}_{\mathrm{x}}$ and $\boldsymbol{J}_{\mathrm{z}} \triangleq \boldsymbol{A}^\dagger\boldsymbol{\Lambda}_{\mathrm{z}}\boldsymbol{A}$ are asymptotically free [17] and from [15, Lemma 3.3.4] we have that[2]

$$\mathrm{R}^K_{\boldsymbol{\Lambda}_{\mathrm{x}}+\boldsymbol{J}_{\mathrm{z}}}(\omega) \simeq \mathrm{R}^K_{\boldsymbol{\Lambda}_{\mathrm{x}}}(\omega) + \mathrm{R}^K_{\boldsymbol{J}_{\mathrm{z}}}(\omega), \quad \Im\omega < 0. \tag{48}$$

Here for a $T \times T$ symmetric matrix $\boldsymbol{X}$ $\mathrm{R}^T_{\boldsymbol{X}}$ denotes the R-transform of the spectrum of $\boldsymbol{X}$, see e.g. [15], and $\simeq$ stands for the large system approximation that turns to an almost surely equality in the large system limit. Furthermore we introduce

$$\mathcal{R}^T_{\boldsymbol{X}}(r) \triangleq \lim_{\omega\to r} \Re\mathrm{R}^T_{\boldsymbol{X}}(\omega), \quad \Im r = 0 \tag{49}$$

whenever the limit exists.

It turns out that by *solely* invoking "additive free convolution", e.g. (48), we can easily solve the complexity issue of the fixed-point identities for $\mathbf{V}_{\mathrm{x}}$ and $\mathbf{V}_{\mathrm{z}}$ which do not require matrix inversion. First we consider the simplification for $\mathbf{V}_{\mathrm{x}}$. To that end let us first define the auxiliary variable

$$q \triangleq \frac{1}{K}\mathrm{tr}\{(\boldsymbol{\Lambda}_{\mathrm{x}} + \boldsymbol{J}_{\mathrm{z}})^{-1}\} = \frac{1}{K}\sum_{k\in\mathcal{K}} \frac{1}{[\Lambda_{\mathrm{x}}]_{kk} + [\mathrm{V}_{\mathrm{x}}]_{kk}}. \tag{50}$$

[2]In fact we can define the R-transform on the negative real line. However this requires the implicit assumption that $\boldsymbol{\Lambda}$ is positive-definite.

Then by invoking (48) we easily obtain that

$$q \simeq \frac{1}{K}\sum_{k\in\mathcal{K}} \frac{1}{[\Lambda_{\mathrm{x}}]_{kk} + \mathcal{R}^K_{\boldsymbol{J}_{\mathrm{z}}}(-q)}. \tag{51}$$

Thereby, we conclude that

$$[\mathrm{V}_{\mathrm{x}}]_{kk} \simeq \mathcal{R}^K_{\boldsymbol{J}_{\mathrm{z}}}(-q), \quad k \in \mathcal{K}. \tag{52}$$

The average of (52) over the random matrix $\boldsymbol{A}$ agrees with [14, Eq. (51)]. Note that the simplification in (52) is still implicit due to the definition of q in (51). Subsequently we present an explicit complexity simplification for $[\mathrm{V}_{\mathrm{x}}]_{kk}$. First we note that (52) states that we can replace all the elements $[\mathrm{V}_{\mathrm{x}}]_{kk}$, $k \in \mathcal{K}$ by a single scalar quantity, say V_{x}. This allows us to write $q \simeq \langle\sigma_{\mathrm{x}}(\boldsymbol{\kappa}_{\mathrm{x}}, \mathrm{V}_{\mathrm{x}}\mathbf{I})\rangle$ with $\boldsymbol{\kappa}_{\mathrm{x}} = \mathrm{V}_{\mathrm{x}}^{-1}\boldsymbol{A}^\dagger\boldsymbol{m} + \hat{\boldsymbol{x}}$. Then, from (52) we write an explicit fixed-point identity for V_{x} as

$$\mathrm{V}_{\mathrm{x}} = \mathcal{R}^K_{\boldsymbol{J}_{\mathrm{z}}}(-\langle\sigma_{\mathrm{x}}(\boldsymbol{\kappa}_{\mathrm{x}}; \mathrm{V}_{\mathrm{x}}\mathbf{I})\rangle). \tag{53}$$

We now address a similar complexity simplification for $[\mathrm{V}_{\mathrm{z}}]_{nn}$ for $n \in \mathcal{N}$. To that end let us introduce an auxiliary $N \times 1$ vector $\tilde{\boldsymbol{\tau}}_{\mathrm{m}}$ whose entries are defined as

$$[\tilde{\tau}_{\mathrm{m}}]_n \triangleq [\Lambda_{\mathrm{z}}]_{nn} - [\Lambda_{\mathrm{z}}]^2_{nn}[A(\Lambda_{\mathrm{x}} + A^\dagger\Lambda_{\mathrm{z}}A)^{-1}A^\dagger]_{nn} \tag{54}$$

$$= [(\Lambda_{\mathrm{z}}^{-1} + A\Lambda_{\mathrm{x}}^{-1}A^\dagger)^{-1}]_{nn}, \tag{55}$$

where (55) follows directly from Woodbury's matrix inversion lemma. Furthermore by making use of (36) for (54) we can write the following fixed-point identity

$$[(\Lambda_{\mathrm{z}}^{-1} + A\Lambda_{\mathrm{x}}^{-1}A^\dagger)^{-1}]_{nn} = [\Lambda_{\mathrm{z}}]_{nn} - \frac{[\Lambda_{\mathrm{z}}]^2_{nn}}{[\Lambda_{\mathrm{z}}]_{nn} + [\mathrm{V}_{\mathrm{z}}]_{nn}} \tag{56}$$

$$= \frac{1}{[\Lambda_{\mathrm{z}}^{-1}]_{nn} + [\mathrm{V}_{\mathrm{z}}^{-1}]_{nn}}. \tag{57}$$

Thus, we can invoke identical arguments on the additive free convolution approximation above for $[\mathrm{V}_{\mathrm{z}}]_{nn}$ as well. Specifically, under Assumption 1-b) and Assumption 2, for large N, K we have

$$[\mathrm{V}_{\mathrm{z}}]_{nn} \simeq \frac{1}{\mathcal{R}^N_{\boldsymbol{J}_{\mathrm{x}}}(-\langle\tilde{\boldsymbol{\tau}}_{\mathrm{m}}\rangle)}, \quad n \in \mathcal{N} \tag{58}$$

with $\boldsymbol{J}_{\mathrm{x}} \triangleq \boldsymbol{A}\boldsymbol{\Lambda}_{\mathrm{x}}^{-1}\boldsymbol{A}^\dagger$. The complexity simplification (58) is still implicit due the definition of $\tilde{\boldsymbol{\tau}}_{\mathrm{m}}$. To present an explicit form of it consider first (56) and (57) such that we can write

$$[\tilde{\tau}_{\mathrm{m}}]_n = [\mathrm{V}_{\mathrm{z}}]_{nn} - \frac{[\mathrm{V}_{\mathrm{z}}]^2_{nn}}{[\mathrm{V}_{\mathrm{z}}]_{nn} + [\mathrm{V}_{\mathrm{z}}]_{nn}} \tag{59}$$

$$= [\mathrm{V}_{\mathrm{z}}]_{nn}\left(1 - [\mathrm{V}_{\mathrm{z}}]_{nn}[\sigma_{\mathrm{z}}(\boldsymbol{\kappa}_{\mathrm{z}}; \mathbf{V}_{\mathrm{z}})]_n\right). \tag{60}$$

On the other hand, (58) implies that we can replace all the elements $[\mathrm{V}_{\mathrm{z}}]_{nn}$, $n \in \mathcal{N}$ by a single scalar quantity, say V_{z}. Now for convenience let us define $N \times 1$ vector $\boldsymbol{\tau}_{\mathrm{m}}$ whose entries are given by

$$[\tau_{\mathrm{m}}]_n \triangleq \mathrm{V}_{\mathrm{z}}\left(1 - \mathrm{V}_{\mathrm{z}}[\sigma_{\mathrm{z}}(\boldsymbol{\kappa}_{\mathrm{z}}; \mathrm{V}_{\mathrm{z}}\mathbf{I})]_n\right), \quad n \in \mathcal{N}. \tag{61}$$

Then following (58) we introduce an explicit fixed-point identity for V_{z} as

$$\mathrm{V}_{\mathrm{z}} = \frac{1}{\mathcal{R}^N_{\boldsymbol{J}_{\mathrm{x}}}(-\langle\boldsymbol{\tau}_{\mathrm{m}}\rangle)}. \tag{62}$$

So far we have shown in (53) and (62) how to bypass the need for matrix inversion to "update" $\mathbf{V}_{\mathrm{x}}$ and $\mathbf{V}_{\mathrm{z}}$, respectively. However this treatment requires solving a highly non-trivial random matrix problem i.e. deriving the closed-form solution for $\mathrm{R}^{K}_{\boldsymbol{J}_{\mathrm{z}}}$ and $\mathrm{R}^{N}_{\boldsymbol{J}_{\mathrm{x}}}$. This is usually, though not always, not possible. On the other hand deriving the solution of e.g $\mathrm{R}^{K}_{\boldsymbol{J}_{\mathrm{z}}}$ in the limiting case is rather simpler. Due to the uniform convergence property of the R-transform [15, Lemma 3.3.4], this approach would allow us to accurately predict, for example $\mathrm{R}^{K}_{\boldsymbol{J}_{\mathrm{z}}}$, for large N, K. This is what we show in the next subsection for the zero-mean iid Gaussian matrix ensemble.

The zero-mean and iid case, i.e. GAMP: In the sequel we provide the explicit solutions for V_{x} and V_{z} when the entries of $\boldsymbol{A}$ are assumed to be iid Gaussian with zero mean and variance $1/N$.

Under Assumption 2 we obtain that

$$\mathrm{R}^{K}_{\boldsymbol{J}_{\mathrm{z}}}(\omega) \simeq \frac{1}{N}\sum_{n\in\mathcal{N}}\frac{1}{[\Lambda_{\mathrm{z}}^{-1}]_{nn}-\omega/\alpha} \tag{63}$$

$$\mathrm{R}^{N}_{\boldsymbol{J}_{\mathrm{x}}}(\omega) \simeq \frac{1}{\alpha K}\sum_{k\in\mathcal{K}}\frac{1}{[\Lambda_{\mathrm{x}}]_{kk}-\omega}. \tag{64}$$

These results can be simply derived by conveniently formulating the well-known Marčhenko-Pastur theorem in terms of the R-transform. While the Marčhenko-Pastur theorem relies on the assumptions that the entries of $\boldsymbol{A}$ are iid (not necessarily Gaussian) with zero mean and $\boldsymbol{A}$ is independent of $\boldsymbol{\Lambda}$ [18], due to the asymptotic freeness, the same holds when the entries are restricted to be Gaussian distributed but without restriction that $\boldsymbol{A}$ and $\boldsymbol{\Lambda}$ are independent. From (63) and (64) we have

$$\mathrm{V}_{\mathrm{x}} \simeq \frac{1}{N}\sum_{n\in\mathcal{N}}\frac{1}{[\Lambda_{\mathrm{z}}^{-1}]_{nn}+\langle\sigma_{\mathrm{x}}(\boldsymbol{\kappa}_{\mathrm{x}};\mathrm{V}_{\mathrm{x}}\mathbf{I})\rangle/\alpha} \tag{65}$$

$$\frac{1}{\mathrm{V}_{\mathrm{z}}} \simeq \frac{1}{\alpha K}\sum_{k\in\mathcal{K}}\frac{1}{[\Lambda_{\mathrm{x}}]_{kk}+\langle\boldsymbol{\tau}_{\mathrm{m}}\rangle}. \tag{66}$$

From these equations one can conclude that

$$\mathrm{V}_{\mathrm{z}} \simeq \frac{\alpha}{\langle\sigma_{\mathrm{x}}(\boldsymbol{\kappa}_{\mathrm{x}};\mathrm{V}_{\mathrm{x}}\mathbf{I})\rangle}, \quad \mathrm{V}_{\mathrm{x}} \simeq \langle\boldsymbol{\tau}_{\mathrm{m}}\rangle. \tag{67}$$

Thereby we recover the fixed points of the GAMP-2nd order updates for the zero-mean iid matrix ensemble as in (15).

IV. Conclusion

For the given zero-mean iid Gaussian matrix ensemble, the fixed points of GAMP "asymptotically" coincide with the stationary points of the Gibbs free energy under first- and second-moment constraints. It turns out that the only critical issue for GAMP is the update rules for "variance" parameters $\mathbf{V}_{\mathrm{x}}$ and $\mathbf{V}_{\mathrm{z}}$. These parameters play a central role. Specifically a crude update rule for a given measurement matrix ensemble would completely spoil the optimality of the algorithm. If for general invariant matrix ensembles, $\mathbf{V}_{\mathrm{x}}$ and $\mathbf{V}_{\mathrm{z}}$ can be updated based on the R-transform formulation in (53) and (62) the algorithm "asymptotically" fulfills the stationary points identities of the Gibbs free energy formulation. Once the closed form expressions of (53) and (62) are obtained, the resulting algorithm includes solely $O(N)$ operations. But the computation of the solutions to these identities is not trivial. Nevertheless it is sometimes doable, e.g. for the random row orthogonal matrix ensembles. Furthermore once either the prior or the likelihood is expressed in terms of a Gaussian function, the R-transform formulation becomes rather trivial. In general updating $\mathbf{V}_{\mathrm{x}}$ and $\mathbf{V}_{\mathrm{z}}$ requires a matrix inversion at each iteration, for example the iteration steps in (43)–(47).

Acknowledgment

This work was supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMmunications NEWCOM♯ (Grant agreement no. 318306).

References

[1] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, pp. 18 914–18 919, September 2009.

[2] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices," *Journal of Statistical Mechanics: Theory and Experiment*, no. P08009, 2012.

[3] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Saint-Petersburg, Russia, July 2011.

[4] J. S. Yedidia, W. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2282–2312, 2005.

[5] E. Riegler, G. E. Kirkelund, C. N. Manchon, M. Badiu, and B. H. Fleury, "Merging belief propagation and the mean field approximation: A free energy approach," *IEEE Transactions on Information Theory*, vol. 59, no. 1, pp. 588–602, 2013.

[6] T. Heskes, M. Opper, W. Wiegerinck, O. Winther, and O. Zoeter, "Approximate inference techniques with expectation constraints," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, 2005.

[7] M. Opper and O. Winther, "Expectation consistent approximate inference," *Journal of Machine Learning Research*, 6 (2005): 2177-2204.

[8] B. Çakmak, O. Winther, and B. H. Fleury, "S-AMP: Approximate message passing for general matrix ensembles," in *Proc. IEEE Information Theory Workshop (ITW)*, Hobart, Tasmanina, Australia, November 2014.

[9] Y. Xu, Y. Kabashima, and L. Zdeborová, "Bayesian signal reconstruction for 1-bit compressed sensing," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2014, 2014.

[10] A. Manoel, F. Krzakala, E. W. Tramel, and L. Zdeborová, "Sparse estimation with the swept approximated message-passing algorithm," *arXiv:1406.4311*, Jun 2014.

[11] S. Rangan, P. Schniter, E. Riegler, A. Fletcher, and V. Cevher, "Fixed points of generalized approximate message passing with arbitrary matrices," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Istanbul, Turkey, July 2013.

[12] M. Opper and O. Winther, "Gaussian processes for classification: Mean-field algorithms," *Neural Computation*, pp. 2655–2684, 2000.

[13] T. P. Minka, "Expectation propagation for approximate Bayesian inference," in *Proc. 17th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2001.

[14] M. Opper and O. Winther, "Adaptive and self-averaging Thouless-Anderson-Palmer mean field theory for probabilistic modeling," *Physical Review E*, vol. 64, pp. 056 131–(1–14), October 2001.

[15] F. Hiai and D. Petz, *The Semicirle Law, Free Random Variables and Entropy*. American Mathematical Society, 2006.

[16] M. Mézard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond*. World Scientific Lecture Notes in Physics-Vol.9, 1987.

[17] B. Collins and P. Sniady, "Integration with respect to the Haar measure on unitary, orthogonal and symplectic group," *Communications in Mathematical Physics*, vol. 264, no. 3, pp. 773–795, 2006.

[18] J. W. Silverstein and Z. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 54, pp. 175–192, 1995.